



III MCSM

Modelagem Computacional no
Sertão Mineiro



PPGMCS
Programa de Pós-Graduação em
Modelagem Computacional e Sistemas

MODELAGEM DE TÓPICOS POR FATORAÇÃO DE MATRIZES NÃO-NEGATIVAS (NMF): MAPEAMENTO CONCEITUAL SOBRE IA E SISTEMAS AGRIVOLTAICOS

Igor Patrick Freitas da Silva¹ - igorfreitaz12@gmail.com

Laís Freitas Coutinho dos Santos² - freitaslais9@gmail.com

Petr Ekel¹ - petr.ekel2709@gmail.com

¹Pontifícia Universidade Católica de Minas Gerais, PUC Minas - Belo Horizonte, MG, Brazil

²Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

Resumo. A rápida evolução da tecnologia fotovoltaica em áreas agrícolas consolidou os sistemas agrivoltaicos, uma combinação de geração fotovoltaica e produção agrícola no mesmo terreno. Paralelamente, a Inteligência Artificial (IA), por meio de vertentes como visão computacional, algoritmos de otimização e de aprendizado de máquina, consolidou-se como um dos pilares da Indústria 4.0 e da agricultura de precisão, ampliando sua aplicação na previsão de geração de energia, no dimensionamento de sistemas e no manejo agrícola. Este trabalho propõe uma abordagem bibliométrica centrada na aplicação da Fatoração de Matrizes Não-Negativas (NMF) para mapear e extrair a estrutura de tópicos latentes desse campo interdisciplinar envolvendo IA e sistemas agrivoltaicos. Foram analisados 160 artigos indexados na base Scopus (2003–2026), processados via pacote `litstudy` (Python). Por meio da decomposição matricial dos termos dos resumos, o algoritmo NMF permitiu agrupar tematicamente os artigos da literatura. Os resultados da modelagem revelaram uma expansão acelerada da área a partir de 2021, com predominância de tópicos associados à previsão de geração de energia e à otimização por aprendizado de máquina. Em contrapartida, observou-se que as investigações sobre a interação direta entre painéis fotovoltaicos e culturas agrícolas ainda são pouco exploradas. Conclui-se que há espaço para uma agenda de pesquisa mais equilibrada entre os componentes energético e agrícola para expansão dos sistemas agrivoltaicos.

Keywords: Sistemas Agrivoltaicos, Inteligência Artificial, Aprendizado de Máquina, Modelagem Computacional, Bibliometria.

1. Introdução

O nexo entre água, energia e alimentos constitui um dos desafios mais críticos da sociedade contemporânea, e é nesse cenário que a fronteira científica da tecnologia fotovoltaica avança

rapidamente, dando origem a conceitos disruptivos como o compartilhamento solar (*solar sharing*), no qual painéis fotovoltaicos e produção agrícola coexistem sobre o mesmo terreno (Dinesh & Pearce, 2016). Paralelamente, a Inteligência Artificial (IA) consolidou-se, em suas diversas ramificações – visão computacional, algoritmos de otimização, redes neurais artificiais e *Big Data* –, como um dos pilares centrais tanto da Indústria 4.0 quanto da Agricultura de Precisão. A confluência dessas duas fronteiras dá origem a um campo de pesquisa emergente: o cruzamento entre a modelagem de dados avançada, fornecida pela IA, e o planejamento espacial e gerencial de usinas agrivoltaicas, que dependem de previsões precisas de radiação, sombreamento e produtividade agrícola para viabilizar sua operação (Weselek et al., 2019).

Entretanto, as fronteiras temáticas desse campo são amplas e encontram-se dispersas de forma fluida pela literatura científica. O grande volume de produção científica sobre IA, de modo geral, dificulta isolar o subconjunto de estudos genuinamente voltado à agrivoltaica. O crescimento acelerado desse campo interdisciplinar gera uma literatura fragmentada entre periódicos de engenharia elétrica, ciência da computação e ciências agrárias, o que dificulta a construção de uma visão global e atualizada do estado da arte.

Diante desse cenário, abordagens tradicionais de revisão sistemática de literatura mostram-se limitadas frente ao volume e à heterogeneidade dos metadados disponíveis, o que reforça a relevância de técnicas de aprendizado de máquina não supervisionado para a extração de padrões ocultos na estrutura conceitual de um campo científico. Entre essas técnicas, a Fatoração de Matrizes Não-Negativas (NMF, do inglês *Non-negative Matrix Factorization*) destaca-se como uma abordagem robusta de modelagem de tópicos, fundamentada em álgebra linear e IA. Em comparação com métodos tradicionais, como a Alocação Latente de Dirichlet (LDA, do inglês *Latent Dirichlet Allocation*), a NMF tende a gerar tópicos mais coerentes, interpretáveis e sem sobreposição negativa entre si a partir de matrizes de termos-documentos de alta dimensionalidade (Choo et al., 2013). Por este motivo, a NMF foi adotada neste trabalho como a técnica central de modelagem de tópicos.

Este trabalho tem como objetivo avaliar o estado da arte das publicações científicas indexadas que correlacionam a Inteligência Artificial ao domínio de sistemas agrivoltaicos, utilizando a NMF como ferramenta central de modelagem de tópicos. Pretende-se, com isso, fornecer um panorama conceitual que sirva tanto como referencial analítico quanto como norteador metodológico para pesquisadores e tomadores de decisão da área.

2. Metodologia

A metodologia do estudo está estruturada em três etapas: (i) definição dos termos de busca e coleta de dados; (ii) processamento dos dados por meio da NMF, utilizando a ferramenta *litstudy*; e (iii) análise e interpretação dos resultados.

A coleta de dados foi realizada em julho de 2026 a partir de uma busca avançada na base de dados *Scopus* da Elsevier, escolhida por sua ampla cobertura de periódicos revisados por pares nas áreas de engenharia, ciências computacionais e ciências agrárias. A *string* de busca combinou termos relativos à energia agrivoltaica com termos relativos à IA, restringindo o resultado a artigos publicados em periódicos (*document type: article*) em língua inglesa. O termo de busca utilizado é detalhado a seguir:

```
TITLE-ABS-KEY("agrivoltaic*" OR "agro-voltaic*" OR  
"agri-photovoltaic*" OR "agro-photovoltaic*" OR  
"solar sharing" OR ("solar energy" AND ("agriculture"  
OR "farming" OR "crop*")) AND TITLE-ABS-KEY
```

```
("artificial intelligence" OR "AI" OR "machine learning"  
OR "deep learning" OR "neural network*" OR  
"reinforcement learning" OR "computer vision" OR  
"optimization algorithm*" OR "big data") AND  
(LIMIT-TO(DOCTYPE,"ar")) AND (LIMIT-TO(LANGUAGE,"English"))
```

O processamento e a análise bibliométrica foram executados em ambiente Python, utilizando a biblioteca *litstudy* (Heldens et al., 2022) para a extração de metadados científicos e cálculo de estatísticas descritivas. Para a análise de agrupamento temático e dos indicadores de estrutura conceitual, construiu-se um *corpus* textual unificado a partir da junção dos títulos, resumos e palavras-chave dos 160 documentos. Adicionalmente, computou-se a frequência dos termos mais relevantes para apoiar a interpretação qualitativa dos dados.

Sobre esse *corpus*, aplicou-se o modelo de agrupamento de tópicos (*topic modeling*) baseado na NMF. A técnica representa o *corpus* como uma matriz termo-documento V não negativa de dimensão $m \times n$, em que m é o número de termos do vocabulário e n o número de documentos. O algoritmo busca aproximar V pelo produto de duas matrizes de posto reduzido, W ($m \times k$) e H ($k \times n$), onde k representa o número de tópicos definidos. O objetivo é minimizar a norma de Frobenius da diferença entre V e WH , sujeito à restrição de não negatividade de ambas as matrizes (Shahnaz et al., 2006), ou seja, a otimização é definida por:

$$\min_{W,H} \|V - WH\|_F^2,$$

onde $W_{ij} > 0$ e $H_{ij} > 0$, para cada i e j . Cada uma das k colunas de W corresponde a um tópico, caracterizado pelo conjunto de termos de maior peso, enquanto cada coluna de H indica o grau de associação de um documento aos k tópicos, permitindo classificá-lo no tópico de maior peso. Por admitir apenas combinações aditivas dos vetores-base, ao contrário de técnicas como a análise de componentes principais (PCA), cujos vetores podem conter entradas negativas, a NMF pode fornecer associações de tópicos coerentes e interpretáveis a partir de *corpora* textuais esparsos e de alta dimensionalidade (Shahnaz et al., 2006).

3. Resultados e discussão

Esta seção apresenta os resultados obtidos da aplicação da metodologia. Inicialmente, apresenta-se a um histograma de distribuição de termos mais frequentes na base de dados analisada. Em seguida, apresenta-se os resultados do agrupamento de tópicos baseado em NMF.

3.1 Análise de distribuição dos tópicos

A frequência relativa dos termos mais recorrentes em todo o volume textual foi calculada e representada no histograma de distribuição apresentado na Figura 1. A análise de frequência de termos do *corpus* (títulos, resumos e palavras-chave) revelou, após o termo estrutural *crop* (proveniente da própria *string* de busca), maior recorrência dos termos *systems*, *power*, *prediction*, *network*, *optimization*, *generation*, *efficiency*, *effective* e *machine*, evidenciando que o núcleo da produção científica estava orientado à modelagem preditiva e à otimização da geração energética, sem se dissociar do componente agrícola.

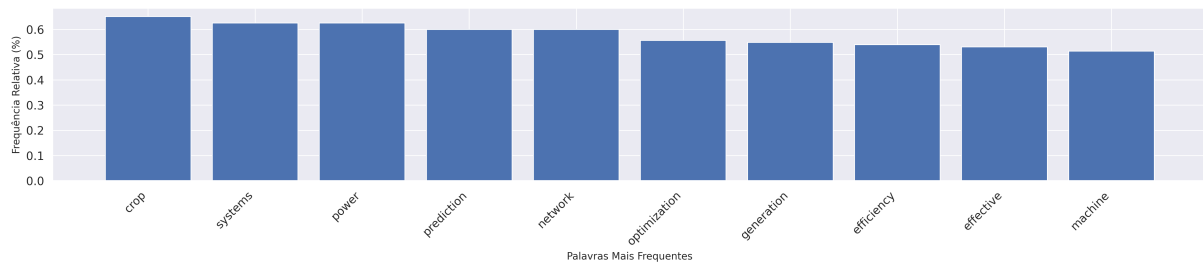


Figura 1: Palavras mais frequentes nos documentos analisados.

3.2 Agrupamento de tópicos

O *corpus* foi segmentado em $k = 10$ tópicos. Assim, a aplicação do modelo baseado em NMF resultou na extração de dez tópicos latentes. Cada agrupamento foi caracterizado pelas palavras de maior peso em sua respectiva coluna da matriz W , visualizadas por meio de nuvens de palavras, as quais estão apresentadas na Figura 2.

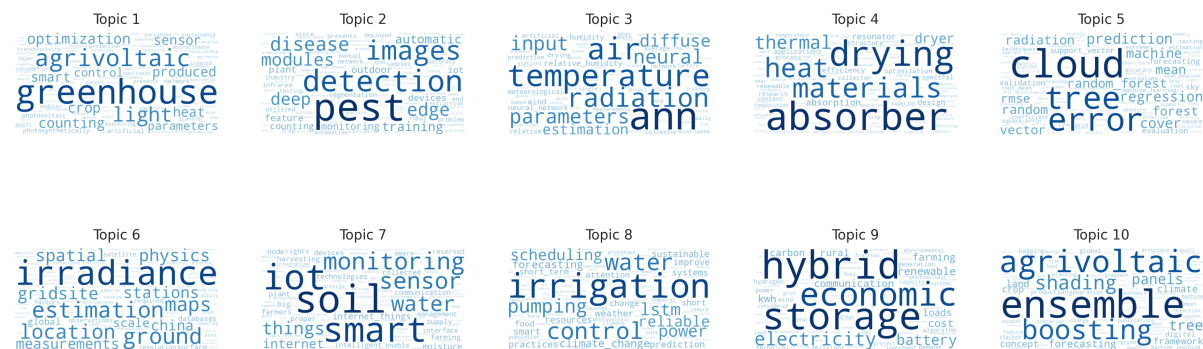


Figura 2: Nuvens de palavras dos dez tópicos identificados pelo modelo de agrupamento temático.

O Tópico 1 inclui termos como *greenhouse*, *agrivoltaic* e *optimization*, agrupando estudos relacionados, por exemplo, à suplementação luminosa artificial em estufas agrivoltaicas (Rampinelli et al., 2024). O Tópico 2 inclui *pest*, *detection*, *images*), sugerindo estudos de monitoramento de fatores bióticos/abióticos e do manejo de precisão sob os módulos, incluindo a identificação automatizada de fitossanidades por visão computacional (Hung et al., 2025). O Tópico 3 traz *ann*, *temperature*, *air*. O Tópico 4 (*absorber*, *drying*, *materials*) e inclui estudo de secagem solar de grãos e classificação de culturas via IoT (Sharma et al., 2022). O Tópico 5 (*cloud*, *tree*, *error*) inclui um estudo sobre predição de radiação solar em regiões semi-áridas Adam and Hamad (2025). O Tópico 6 (*irradiance*, *estimation*, *ground*) inclui um estudo sobre radiação e irradiância no solo com *deep learning* (Kurumundayil et al., 2025). Os Tópicos 7 (*soil*, *iot*, *smart*) e 8 (*irrigation*, *control*, *water*) trazem estudos que aplicam IoT no controle da água e do solo (Banluesapy et al., 2025; Kee et al., 2025). Os Tópicos 9 (*hybrid*, *storage*, *economic*) e 10 (*ensemble*, *agrivoltaic*, *boosting*) incluem estudos de geração híbrida e *ensemble learning* para armazenamento e viabilidade econômica (Datta et al., 2026).

Por fim, a relação espacial entre os dez tópicos foi examinada por meio de uma projeção bidimensional dos documentos, na qual cada ponto representa um artigo colorido conforme seu tópico dominante. A projeção dos documentos na Figura 3 permite visualizar a proximidade

ou o distanciamento temático entre os grupos identificados. A distribuição espacial desses agrupamentos na projeção bidimensional permite confirmar a natureza interdisciplinar do campo, revelando frentes de pesquisa entre agricultura de precisão, engenharia, sistemas fotovoltaicos e viabilidade econômica.

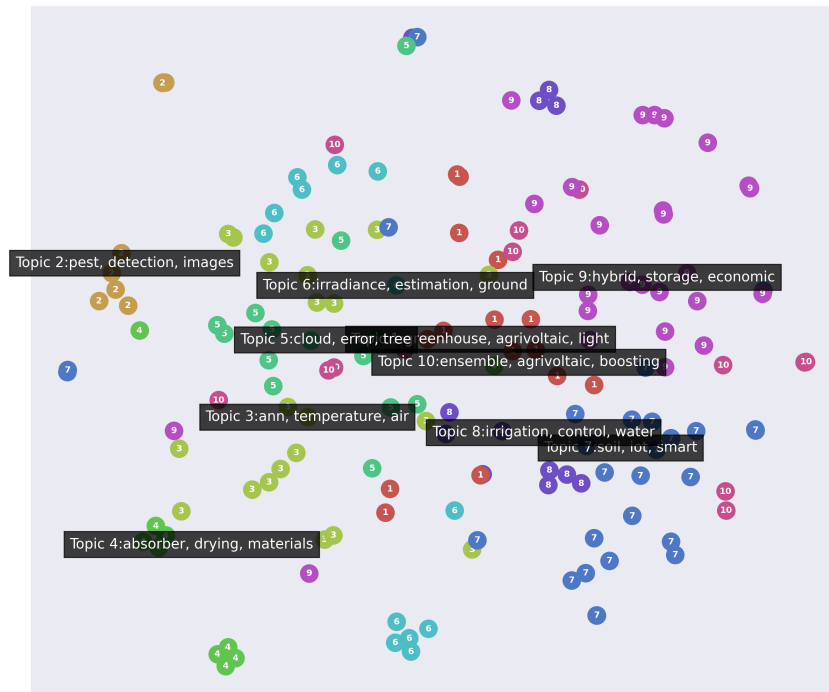


Figura 3: Projeção bidimensional dos documentos por tópico dominante.

4. Considerações Finais

Este trabalho apresentou um mapeamento bibliométrico da produção científica envolvendo sistemas agrivoltaicos e IA, analisando uma base de dados com 160 artigos indexados no *Scopus* por meio de um método baseado em NMF. Com base nos resultados obtidos, verifica-se que a pesquisa em sistemas agrivoltaicos pode ser estruturada em três eixos. O primeiro, representado principalmente pelos Tópicos 1 a 4, concentra-se nos aspectos físicos e ambientais do cultivo. O segundo eixo, relacionado aos Tópicos 5 a 8, evidencia o papel crescente das técnicas computacionais e de IoT no monitoramento e controle de variáveis agronômicas, como a estimativa de irradiância no solo, a avaliação de erros em modelos preditivos e o gerenciamento da irrigação e do solo. Por fim, o terceiro eixo, composto pelos Tópicos 9 e 10, direciona-se a questões econômicas, abordando sistemas híbridos de geração e armazenamento de energia e a viabilidade econômica desses sistemas. Do ponto de vista prático, os resultados obtidos oferecem um panorama estruturado dos temas consolidados e das lacunas existentes em IA para sistemas agrivoltaicos. Esse estudo fornece subsídios para a tomada de decisão em projetos agrivoltaicos, evidenciando que soluções baseadas em IA podem ser exploradas para aprimorar a sinergia entre geração fotovoltaica e produtividade agrícola.

Agradecimentos

Os autores agradecem à PUC-Minas (Processo 2026/35102), ao CNPq (Processos nº 407180/2023-0 e 300141/2026-2), à FAPEMIG (Processo APQ-04802-24 e APD-00860-25), à FIP 2025/32418 e à FIP 2026/35102.

Declaração de Uso de Inteligência Artificial

Em conformidade com a Portaria CNPq nº 2.664/2026, os autores declaram o uso da ferramenta de Inteligência Artificial Generativa Claude (Anthropic) nas seguintes etapas deste trabalho: (i) verificação e refinamento da *string* de busca bibliográfica utilizada na base Scopus; (ii) checagem da correspondência entre os artigos citados como exemplos dos tópicos identificados pelo modelo NMF e o conteúdo de seus resumos originais; (iii) apoio à estruturação e à revisão textual das seções deste artigo. Todo o conteúdo produzido com apoio da ferramenta foi integralmente revisado pelos autores, que assumem total responsabilidade pela originalidade e veracidade do texto final.

Referências

- Adam, A.M., Hamad, A. (2025), “Solar radiation prediction in semi-arid regions: A machine learning approach and comprehensive evaluation in Gadarif, Sudan”, *Italian Journal of Agrometeorology*, vol. 1, pp. 51–67.
- Banluesapy, S., Ketcham, M., Rattanasiriwongwut, M. (2025), “AI-Augmented Smart Irrigation System Using IoT and Solar Power for Sustainable Water and Energy Management”, *Energy Engineering*.
- Choo, J., Lee, C., Reddy, C.K., Park, H. (2013), “UTOPIAN: User-driven topic modeling based on interactive nonnegative matrix factorization”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 1992–2001.
- Datta, L., Kumar, A., Rizwan, M. (2026), “Forecasting agrivoltaic power using a hybrid model accounting for crop and ground effects”, *Electrical Engineering*.
- Dinesh, H., Pearce, J.M. (2016), “The potential of agrivoltaic systems”, *Renewable and Sustainable Energy Reviews*, 54, 299–308.
- Heldens, S., Sclocco, A., Dreuning, H., Van Werkhoven, B., Hijma, P., Maassen, J., Van Nieuwpoort, R.V. (2022), “litstudy: A Python package for literature reviews”, *SoftwareX*, 20, 101207.
- Hung, C.-W., Wang, C.-C., Liao, Z.-J., Su, Y.-H., Liu, C.-L. (2025), “Intelligent Battery-Designed System for Edge-Computing-Based Farmland Pest Monitoring System”, *Electronics (Switzerland)*.
- Kaplan, Y.A., Tolun, G.G., Kaplan, A.G. (2026), “Developing machine learning-based quadratic Angström models for solar radiation prediction”, *Journal of Atmospheric and Solar-Terrestrial Physics*.
- Kee, K.-K.; Rashidi, R.; Kee, O.K.-H.; Han, A.B.; Patrick, I.Z.; Bawen, L.M. Context-aware self-powered intelligent soil monitoring system for precise agriculture. *International Journal of Electrical and Computer Engineering*, v. 15, n. 1, p. 1123–1131, 2025. DOI: 10.11591/ijece.v15i1.pp1123-1131
- Kurumundayil, L., Burkhardt, D., Gföllner, L., Rupitsch, S.J., Preu, R., Berwind, M., Demant, M. (2025), “Fast ground irradiance computations for agrivoltaics via physics-informed deep learning models”, *Communications Engineering*.
- Rampinelli, G.A.; Marcelino, R.; Possenti Damasceno, J.; Caroline Stalter, C.; Bouchardet, A.T.R.; Mohr, G.; Guber, V. Development of artificial lighting system for light supplementation in smart greenhouses with agrivoltaic systems. *Renewable Energy*, v. 231, art. 120914, 2024. DOI: 10.1016/j.renene.2024.120914
- Shahnaz, F., Berry, M.W., Pauca, V.P., Plemmons, R.J. (2006), “Document clustering using nonnegative matrix factorization”, *Information Processing & Management*, 42(2), 373–386.
- Sharma, B.B., Gupta, G., Vaidya, P., Basheer, S., Memon, F.H., Thakur, R.N. (2022), “Internet of Things-Based Crop Classification Model Using Deep Learning for Indirect Solar Drying”, *Wireless Communications and Mobile Computing*.
- Weselek, A., Ehmann, A., Zikeli, S., Lewandowski, I., Schindele, S., Högy, P. (2019), “Agrophotovoltaic systems: applications, challenges, and opportunities. A review”, *Agronomy for Sustainable Development*, 39, 35.